



# A Study on Faceted Search for Semi-structured Data

著者	駒水 孝裕
発行年	2015
その他のタイトル	半構造データに対するファセット検索に関する研究
学位授与大学	筑波大学 (University of Tsukuba)
学位授与年度	2014
報告番号	12102甲第7280号
URL	<a href="http://hdl.handle.net/2241/00125879">http://hdl.handle.net/2241/00125879</a>

氏 名 (本籍)	駒水 孝裕		
学 位 の 種 類	博 士 ( 工 学 )		
学 位 記 番 号	博 甲 第	7280	号
学位授与年月日	平成 27 年 3 月 25 日		
学位授与の要件	学位規則第 4 条第 1 項該当		
審 査 研 究 科	システム情報工学研究科		
学位論文題目	A Study on Faceted Search for Semi-structured Data (半構造データに対するファセット検索に関する研究)		
主 査	筑波大学 教授	理学博士	北川 博之
副 査	筑波大学 教授	博士 (工学)	山本 幹雄
副 査	筑波大学 教授	博士 (工学)	佐藤 哲司
	(図書館情報メディア研究科)		
副 査	筑波大学 准教授	博士 (工学)	天笠 俊之
副 査	筑波大学 准教授	博士 (理学)	南出 靖彦

## 論 文 の 要 旨

スキーマ情報等により厳格に構造が規定されない半構造データは、その柔軟性ゆえに、複雑な実世界の実体やイベントなどの記述にしばしば利用される。半構造データに対する検索は、半構造データを利活用する上で重要である。これまで、半構造データを対象とした問合せ言語による検索やキーワード検索についての研究が行われており、これらは、検索要求が明確な場合に利用することができる。しかし、問合せ結果を得ながら対話的に絞り込みを行うような探索的な検索には、必ずしも適切ではない。このような探索的な検索を支援する技術としてファセット検索があるが、従来の主な対象は構造が明確な属性付きオブジェクトが主であり、半構造データを対象としたファセット検索については、一部先行研究があるとはいえ、一般的な枠組みの提案はこれまで行われてこなかった。本論文では、半構造データに対するファセット検索の基本フレームワークを提案すると共に、より適切なオブジェクト、ファセットの抽出手法を提案し、実験評価により有効性を確認している。

ファセット検索は、属性付きのオブジェクトを検索対象とし、属性をオブジェクト集合の特徴を表すファセットとして提示する。検索者はファセットからオブジェクト集合の特徴を読み取り、更なる絞り込みを行うヒントを得る。

本論文では、ファセット検索を半構造データに適用する際の課題について、4つのポイントから解決法を示している。ファセット検索を半構造データの検索に適用するためには、データ内のオブジェクトとファセットを予め抽出する必要があるが、半構造データでは何をオブジェクトとし何をファセットとするかが明示的には明らかでない。本論文では、まず XML 等の半構造木データを対象に、繰り返し出現する部分木をオブジェクト、その子孫リーフをファセットとするファセット検索のフレームワークを提案し、DBLP データ等を対象としたシステムを構築し、探索的検索における有効性を示し

ている．次に，グラフデータを対象に，頻出部分グラフをオブジェクト，その周辺ノードをファセットとするフレームワークを提案している．さらに，XML 等の半構造木データにおけるより適切なオブジェクト抽出の実現のため，出現頻度に基づく方法，意味に基づく方法，両者を組合せた方法を提案し，有効性の評価を行っている．出現頻度に基づく方法では，出現回数の閾値を設定し，それを超えるような頻出のものを抽出する．意味に基づく方法では，Wikipedia や WordNet といった外部リソースを活用し，実体として認識されやすいものを抽出する．最後に，XML 等の半構造木データにおけるテキストデータの扱いについて検討を行っている．具体的にはテキストデータ全体をファセット値とするのではなく，その中から抽出した単語をファセット値として用いる方式を提案すると共に，単語の共起度から概念階層を構築し，利用者へファセット値を提示する際にその概念階層を利用する方式を提案し，実験評価によって有効性を示している．

## 審 査 の 要 旨

### 【批評】

本研究は，木やグラフに代表される半構造データに対する探索的検索を支援するためのファセット検索の枠組みについて，複数の視点から提案と考察を行っている．ファセット検索を半構造データの検索に適用する際に，検索対象オブジェクトとファセットを抽出する必要があるという課題に対しては，頻出する部分木や部分グラフをオブジェクトとするという，汎用性の高いアプローチを示している．また，その際に外部リソースの活用についても検討を行っている．さらに，テキストデータを含む半構造データに対するファセット検索の利便性の向上のために，テキスト中の単語を抽出すると共にその概念階層を利用するという手法を示している．これらを通して，これまで研究が十分行われてこなかった半構造データに対するファセット検索に対して体系的な知見を与えた点で，新規性と情報工学上の貢献が認められる．また，実験システム構築や評価実験を通して，提案手法の有効性を示している点も評価できる．今後，ファセット検索におけるファセットのランキング，より複雑な構造をもった半構造データの扱い等，残された課題に対する検討が進められることが期待される．

### 【最終試験の結果】

平成 27 年 2 月 4 日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

### 【結論】

上記の学位論文審査ならびに最終試験の結果に基づき、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。

